

Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Correction of incoherent conditional probability assessments

A. Capotorti *, G. Regoli, F. Vattari

Dip. Matematica e Informatica, Università di Perugia, 06123 Perugia, Italy

ARTICLE INFO

Article history:

Received 19 June 2009

Received in revised form 2 February 2010

Accepted 5 February 2010

Available online 19 February 2010

Keywords:

Conditional probability coherence

Divergence

Scoring rules

ABSTRACT

In this paper we deep in the formal properties of an already stated discrepancy measure between a conditional assessment and the class of unconditional probability distributions compatible with the assessment domain.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

In this paper we deal with incoherent partial conditional probabilities assessments. Such kind of evaluations arise because often it is natural to give evaluations of probability only on relevant events, that are judged under specific circumstances. And it can happen that the numerical values do not fit well with each other, especially when information comes from different sources.

Inconsistency, if not adjusted, can be dangerous. In fact, often the assessment is intended to be used for inference purposes, i.e. to see how a further (conditional) event can be evaluated consistently with the initial assessment. Of course, the inferential results are meaningful only if the prior information encompassed in the initial assessment is coherent by itself.

Hence it is quite natural to search for a coherent assessment on the same domain that will preserve the opinion expressed by the initial assessment as much as possible, without introducing exogenous information. This goal is obtained by minimizing some kind of distance among partial conditional assessments.

Distances and pseudo-distances among probability distributions are usually measured through divergencies (e.g. Euclidean distance, Kulback–Leibler divergence, Csiszár f -divergences, etc.). Some of them can be applied only among unconditional full probability distributions; others could be applied to our context of partial conditional assessments (see for example [10,15]), but do not have a fully convincing probabilistic justification, being purely geometrical tools. Hence, for our purpose, in this paper we introduce an index of “discrepancy” among partial conditional probability assessments which is derived by a particular scoring rule. Such a scoring rule is inspired by the one introduced by Lad in [12] for unconditional probability distributions, and adapted here to partial conditional frameworks.

2. A short description of the problem

We briefly describe the problem.

* Corresponding author. Tel.: +39 0755855011; fax: +39 0755855024.

E-mail addresses: capot@dmf.unipg.it (A. Capotorti), regoli@dmf.unipg.it (G. Regoli), francesca.vattari@dmf.unipg.it (F. Vattari).

A field expert, in the sequel named the “assessor”, elicits a finite family of conditional events $\mathcal{E} = [E_1|H_1, \dots, E_n|H_n]$ as domain of his/her evaluations. The events E_i 's usually represent the situations under consideration, while the H_i 's usually represent the different contexts, or scenarios, under which the E_i 's are evaluated.

The assessor is supposed to elicit numerical values $\mathbf{p} = (p_1, \dots, p_n)$ thought as his/her honest evaluation of the probabilities $P(E_i|H_i)$, $i = 1, \dots, n$. The problem consists to adjust such an evaluation when it turns out to be incoherent, i.e. incompatible with any probability distribution.

2.1. Preliminaries

The basic events $E_1, \dots, E_n, H_1, \dots, H_n$ can be endowed with logical constraints, that represent dependencies among particular configurations of them (e.g. incompatibilities, implications, partial or total coincidences, etc.).

In the following $E_i H_i$ will denote the logical connection “ E_i and H_i ”, E_i^c will indicate “not E_i ” and the event $H^0 = \bigvee_{i=1}^n H_i$ will represent the whole set of contexts.

For the sake of simplicity we skip from the present job the two extreme situations of incompatibility between situation and scenario or of inclusion of the scenario in the situation, i.e.

$$E_i H_i = \phi \quad \text{or} \quad H_i \subseteq E_i. \quad (1)$$

In fact, in these two cases the probability values are compulsorily determined by coherence

$$E_i H_i = \phi \Rightarrow P(E_i|H_i) = 0; \quad H_i \subseteq E_i \Rightarrow P(E_i|H_i) = 1, \quad (2)$$

and any violation of them can be trivially adjusted.

Starting with the basic events $E_1, \dots, E_n, H_1, \dots, H_n$, it is possible to span a sample space $\Omega = \{\omega_1, \dots, \omega_k\}$, where ω_j represents a generic atom that is the minimal element in the algebra generated by E_i, H_i . Note that the sample space Ω , together with H^0 , are not part of the assessment but only auxiliary tools.

Every probability distribution $\alpha : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$ corresponds to a nonnegative vector $\alpha = [\alpha_1, \dots, \alpha_k]$, with $\alpha_j = \alpha(\omega_j)$, then for every event E it results $\alpha(E) = \sum_{\omega_j \in E} \alpha_j$.

We need to introduce a nested hierarchy among sets of probability distributions:

- let $\mathcal{A} = \{\alpha = [\alpha_1, \dots, \alpha_k], \sum_1^k \alpha_i = 1, \alpha_j \geq 0, j = 1, \dots, k\}$ represents the whole set of probability distributions on Ω ;
- let $\mathcal{A}_0 = \{\alpha \in \mathcal{A} | \alpha(H^0) = 1\}$ be the subset of probability distributions on Ω that concentrate all the probability mass on the contemplated scenarios;
- let $\mathcal{A}_1 = \{\alpha \in \mathcal{A}_0 | \alpha(H_i) > 0, i = 1, \dots, n\}$ be the subset of probability distributions on Ω that give positive probability to every scenario;
- let $\mathcal{A}_2 = \{\alpha \in \mathcal{A}_1 | 0 < \alpha(E_i H_i) < \alpha(H_i), i = 1, \dots, n\}$ be the subset of probability distributions that avoid boundary values $\{0, 1\}$ for the conditional probabilities.

It is easy to see that the sets \mathcal{A}_i are convex sets and \mathcal{A}_0 is the closure of \mathcal{A}_2 (and \mathcal{A}_1) in the usual topology.

Note that in conditional frameworks the focusing on \mathcal{A}_0 is commonly done to avoid unpleasant consequences. See Walley [16] about Avoiding Uniform Loss assessments or Holzer [9] about the Principle of Conditional Coherence.

2.2. Coherence

As already mentioned, we focus our attention on inconsistent assessments \mathbf{p} . Consistency for partial assessments can be reduced to the compatibility with a well established mathematical model. For conditional probabilities the reference models are the so called *full conditional probabilities*, as introduced by Dubins [8] and in line also with De Finetti [7], Krauss [11] and Rényi [14] thoughts. Full conditional probabilities are characterized by the following set of axioms:

Definition 1. Given a Boolean algebra \mathcal{B} , a *full conditional probability* on $\mathcal{B} \times \mathcal{B}^0$ ($\mathcal{B}^0 = \mathcal{B} \setminus \{\emptyset\}$) is a function $P : \mathcal{B} \times \mathcal{B}^0 \rightarrow [0, 1]$ such that

- $P(\cdot|H)$ is a finitely additive probability on \mathcal{B} for any given H in \mathcal{B}^0 ;
- $P(H|H) = 1$ for all $H \in \mathcal{B}^0$;
- $P(A|C) = P(A|B)P(B|C)$ for every $A \in \mathcal{B}$, $B, C \in \mathcal{B}^0$, with $A \subseteq B \subseteq C$.

Note that, whenever (i) and (ii) are satisfied, condition (iii) is equivalent to

$$(iii') \quad P(AB|C) = P(B|C)P(A|BC) \quad \text{for every } A, B \in \mathcal{B}, C \in \mathcal{B}^0, BC \in \mathcal{B}^0.$$

The pairs $(A|H) \in \mathcal{B} \times \mathcal{B}^0$ are called conditional events.

Consequently we have:

Definition 2. If $\mathcal{E} = [E_1|H_1, \dots, E_n|H_n]$ is an arbitrary set of conditional events, an assessment $P(\cdot|\cdot)$ on \mathcal{E} is said to be *coherent* if there exists a *full conditional probability* $P'(\cdot|\cdot)$ defined on $\mathcal{P}(\Omega) \times \mathcal{P}(\Omega)^0$ (with $\mathcal{P}(\Omega)$ the power set of Ω) which agrees with $P(\cdot|\cdot)$ on \mathcal{E} .

Every probability distribution $\alpha \in \mathcal{A}_1$ generates a coherent assessments \mathbf{q}_α on \mathcal{E} through the usual formula

$$q_{\alpha_i} = \frac{\sum_{\omega_j \subseteq E_i H_i} \alpha_j}{\sum_{\omega_j \subseteq H_i} \alpha_j} \quad \forall i = 1, \dots, n. \quad (3)$$

Note that \mathbf{q}_α is a continuous function of α when $\alpha \in \mathcal{A}_1$.

When $\alpha \in \mathcal{A}_0$, previous formula (3) defines \mathbf{q}_α only on

$$\mathcal{E}_\alpha := \{E_i|H_i \in \mathcal{E}, \alpha(H_i) > 0\}. \quad (4)$$

Coherence of \mathbf{q}_α is guaranteed by the theorem of Coletti [5] and Coletti and Scozzafava [6].

3. A discrepancy measure

Associated to any (coherent or not) assessment $\mathbf{p} \in (0, 1)^n$ over $\mathcal{E} = [E_1|H_1, \dots, E_n|H_n]$ we can introduce a scoring rule¹

$$S(\mathbf{p}) := \sum_{i=1}^n |E_i H_i| \ln p_i + \sum_{i=1}^n |E_i^c H_i| \ln(1 - p_i), \quad (5)$$

where $|\cdot|$ is the indicator function of unconditional events.

Such score $S(\mathbf{p})$ is an “adaptation” of the “proper scoring rule” for probability distributions proposed by Lad in [12, p. 355]. We have extended it to partial and conditional probability assessments. The motivation of such a score is that, for a conditional event $E_i|H_i$ which is a three-valued logical entity partitioning Ω in three parts (the atoms satisfying $E_i H_i$ and thus verifying the conditional, those satisfying $E_i^c H_i$, thus falsifying the conditional, and those not fulfilling the context H_i , to which the conditional may not be applied at all) the assessor “loses less” the higher are the probabilities assessed for events that are verified, and at the same time, the lower are the probabilities assessed for those that are not verified. The values assessed on events that turn out to be undetermined do not influence the score. In fact the realization of the random value $S(\mathbf{p})$ when the atom ω_j occurs is

$$S_j(\mathbf{p}) = \sum_{E_i H_i \supseteq \omega_j} \ln p_i + \sum_{E_i^c H_i \supseteq \omega_j} \ln(1 - p_i). \quad (6)$$

We can introduce the “discrepancy” between an assessment \mathbf{p} over \mathcal{E} and a distribution $\alpha \in \mathcal{A}_2$, with respect to its conditional coherent assessment \mathbf{q}_α , as

$$\Delta(\mathbf{p}, \alpha) := E_\alpha(S(\mathbf{q}_\alpha) - S(\mathbf{p})) = \sum_{j=1}^k \alpha_j [S_j(\mathbf{q}_\alpha) - S_j(\mathbf{p})]. \quad (7)$$

It is easy to see that

$$\Delta(\mathbf{p}, \alpha) = \sum_{i=1}^n \alpha(E_i H_i) \ln \left(\frac{q_i}{p_i} \right) + \alpha(E_i^c H_i) \ln \left(\frac{1 - q_i}{1 - p_i} \right) = \quad (8)$$

$$= \sum_{i=1}^n \alpha(H_i) \left(q_i \ln \left(\frac{q_i}{p_i} \right) + (1 - q_i) \ln \left(\frac{1 - q_i}{1 - p_i} \right) \right). \quad (9)$$

The restriction to the distributions α in \mathcal{A}_2 is because only there the scoring rule $S(\mathbf{q}_\alpha)$ is properly defined. Anyhow it is possible to extend by continuity $\Delta(\mathbf{p}, \alpha)$ to any distribution α in \mathcal{A}_0 by defining

$$\Delta(\mathbf{p}, \alpha) := \sum_{i: \alpha(H_i) > 0} \alpha(H_i) \left(q_i \ln \left(\frac{q_i}{p_i} \right) + (1 - q_i) \ln \left(\frac{1 - q_i}{1 - p_i} \right) \right), \quad (10)$$

adopting the usual convention $0 \ln(0) = 0$.

3.1. Formal properties of $\Delta(\mathbf{p}, \alpha)$

In the sequel we will use the following two properties:

$$t \ln t - t + 1 \geq 0 \quad \forall t \geq 0, \quad (11)$$

$$\ln t - t + 1 \leq 0 \quad \forall t > 0. \quad (12)$$

¹ Boundary values 0 or 1 for the assessed probabilities are avoided to skip technical drawbacks in the definition of the scoring rule. This is anyhow consistent with our choice of avoiding trivial inconsistencies from the beginning (see (1) and (2), and the associated comments).

In fact, the line $y = t - 1$ is tangent both to the strictly convex function $t \ln t$ and to the strictly concave function $\ln t$ at the point $t = 1$. Moreover in (11) and (12) the equality holds if and only if $t = 1$.

The discrepancy $\Delta(\mathbf{p}, \alpha)$ behaves like other common divergences. In fact the following theorem holds.

Theorem 1. Let $\Delta(\mathbf{p}, \alpha)$ be defined as in (10). Then $\Delta(\mathbf{p}, \alpha)$ is continuous in \mathcal{A}_0 . Moreover

- (i) $\Delta(\mathbf{p}, \alpha) \geq 0 \quad \forall \alpha \in \mathcal{A}_0$;
- (ii) for any $\alpha \in \mathcal{A}_1$, $\Delta(\mathbf{p}, \alpha) = 0$ iff $\mathbf{p} \equiv \mathbf{q}_\alpha$.

Proof. The continuity on \mathcal{A}_2 is trivial. The continuity on \mathcal{A}_0 depends on the fact that

$$q \ln \frac{q}{p} + (1 - q) \ln \left(\frac{1 - q}{1 - p} \right),$$

is bounded for $q \in (0, 1)$ and hence every term in

$$\sum_{i: \alpha(H_i) > 0} \alpha(H_i) \left(q_i \ln \left(\frac{q_i}{p_i} \right) + (1 - q_i) \ln \left(\frac{1 - q_i}{1 - p_i} \right) \right),$$

is continuous even when $\alpha(H_i)$ tends to 0. To prove (i) it suffices to prove that the function

$$f(x, y) := x \ln \frac{x}{y} + (1 - x) \ln \left(\frac{1 - x}{1 - y} \right), \quad (13)$$

is nonnegative for any $x \in [0, 1]$ and $y \in (0, 1)$. But (13) can be rewritten as

$$f(x, y) := x \ln \frac{x}{y} - x + y + (1 - x) \ln \left(\frac{1 - x}{1 - y} \right) - (1 - x) + (1 - y),$$

so that, letting $t := x/y$ (or $t := (1 - x)/(1 - y)$), we have that $f(x, y)$ is the sum of two terms like

$$t \ln t - t + 1,$$

which are nonnegatives by (11).

(ii) If α lies in \mathcal{A}_1 we have $\alpha(H_i) > 0$, $i = 1, \dots, n$ so that $\Delta(\mathbf{p}, \alpha)$ results a sum of nonnegative terms. Hence

$$\Delta(\mathbf{p}, \alpha) = 0 \iff q_i \ln \left(\frac{p_i}{q_i} \right) + (1 - q_i) \ln \left(\frac{1 - p_i}{1 - q_i} \right) = 0 \quad \forall i = 1, \dots, n.$$

Since the function $t \ln t - t + 1$ vanishes only in $t = 1$ we have

$$\begin{aligned} \Delta(\mathbf{p}, \alpha) = 0 &\iff \frac{p_i}{q_i} = 1 \quad \forall i = 1, \dots, n, \\ &\iff \mathbf{p} \equiv \mathbf{q}_\alpha. \quad \square \end{aligned}$$

Note that (ii) in Theorem 1 can be easily generalized for $\alpha \in \mathcal{A}_0$. In fact since in (10) there are only terms with $\alpha(H_i) > 0$, we have

$$\Delta(\mathbf{p}, \alpha) = 0 \iff P(E_i | H_i) = q_{x_i} \quad \forall E_i | H_i \in \mathcal{E}_\alpha,$$

where \mathcal{E}_α is defined as in (4).

Now we can assert something about the convexity of our discrepancy and about its minimum value.

Theorem 2. Let \mathbf{p} be an assessment on $\mathcal{E} = [E_1 | H_1, \dots, E_n | H_n]$ and $\Delta(\mathbf{p}, \alpha)$ defined as in (10), with $\alpha \in \mathcal{A}_0$. Then $\Delta(\mathbf{p}, \alpha)$ is a convex function with respect to α .

Proof. We first prove that $\Delta(\mathbf{p}, \alpha)$ is convex in \mathcal{A}_2 .

The expression (8) of $\Delta(\mathbf{p}, \alpha)$ can be written as

$$\Delta(\mathbf{p}, \alpha) = \sum_{i=1}^n \left[\alpha(E_i H_i) \left(\ln \frac{\alpha(E_i H_i)}{\alpha(E_i H_i) + \alpha(E_i^c H_i)} - \ln p_i \right) + \alpha(E_i^c H_i) \left(\ln \frac{\alpha(E_i^c H_i)}{\alpha(E_i H_i) + \alpha(E_i^c H_i)} - \ln(1 - p_i) \right) \right]. \quad (14)$$

Note that by definition

$$\alpha(E_i H_i) = \sum_{\omega_j \subseteq E_i H_i} \alpha_j, \quad i = 1, \dots, n,$$

and then it will suffice to prove the convexity of Δ with respect to $\alpha(E_i H_i)$ and $\alpha(E_i^c H_i)$, $i = 1, \dots, n$.

Letting $x := \alpha(E_i H_i)$ and $y = \alpha(E_i^c H_i)$ we have that $x, y \in (0, 1)$ because α is in \mathcal{A}_2 . Since Δ is the sum of $2n$ terms like

$$f(x, y) := x \ln \frac{x}{x+y} - x \ln p,$$

in order to prove the convexity of Δ it is sufficient to prove that

$$\nabla f(x_0, y_0)(x - x_0, y - y_0) + f(x_0, y_0) \leq f(x, y).$$

It is easy to see that

$$\nabla f(x_0, y_0) = \left(\ln \frac{x_0}{x_0 + y_0} + \frac{y_0}{x_0 + y_0} - \ln p, -\frac{x_0}{x_0 + y_0} \right),$$

and then $f(x, y)$ is convex if and only if

$$x \ln \left(\frac{x_0(x+y)}{x(x_0+y_0)} \right) \leq \frac{x_0 y - x y_0}{x_0 + y_0}.$$

Since $x > 0$, last inequality reduces to

$$\ln \left(\frac{x_0(x+y)}{x(x_0+y_0)} \right) \leq \frac{x_0(x+y)}{x(x_0+y_0)} - 1,$$

that holds by (12) with $t := x_0(x+y)/x(x_0+y_0)$.

Finally, since \mathcal{A}_0 is the closure of \mathcal{A}_2 and $\Delta(\mathbf{p}, \alpha)$ is continuous in \mathcal{A}_0 , then $\Delta(\mathbf{p}, \alpha)$ is convex on \mathcal{A}_0 . \square

Theorem 3. Let \mathbf{p} be an assessment on $\mathcal{E} = [E_1|H_1, \dots, E_n|H_n]$, then $\Delta(\mathbf{p}, \alpha)$ admits a minimum on \mathcal{A}_0 . Moreover, if such minimum is attained only on \mathcal{A}_1 , then there is a unique coherent assessment \mathbf{q}_α on $\mathcal{E} = [E_1|H_1, \dots, E_n|H_n]$ such that $\Delta(\mathbf{p}, \alpha)$ is minimum.

Proof. Since $\Delta(\mathbf{p}, \cdot)$ is convex in \mathcal{A}_0 that is a convex and closed set, $\Delta(\mathbf{p}, \cdot)$ admits minimum on \mathcal{A}_0 .

Anyhow such minimum could not be unique, i.e. there could be a (convex) set of probability distributions each of them giving the same minimum value $\min_{\alpha \in \mathcal{A}_0} \Delta(\mathbf{p}, \alpha)$.

If the minimum of $\Delta(\mathbf{p}, \cdot)$ is attained only on \mathcal{A}_1 we can prove that if $\alpha, \alpha^0 \in \mathcal{A}_1$ minimize $\Delta(\mathbf{p}, \cdot)$ then $\mathbf{q}_\alpha \equiv \mathbf{q}_{\alpha^0}$.

At first, suppose that

$$(\mathbf{q}_\alpha)_i = 0 \iff (\mathbf{q}_{\alpha^0})_i = 0 \quad \forall i = 1, \dots, n, \quad (15)$$

$$(\mathbf{q}_\alpha)_i = 1 \iff (\mathbf{q}_{\alpha^0})_i = 1 \quad \forall i = 1, \dots, n. \quad (16)$$

So letting

$$x_i := \alpha(E_i H_i), \quad y_i := \alpha(E_i^c H_i), \quad i = 1, \dots, n, \quad (17)$$

$$x_i^0 := \alpha^0(E_i H_i), \quad y_i^0 := \alpha^0(E_i^c H_i), \quad i = 1, \dots, n, \quad (18)$$

we have

$$x_i = 0 \iff x_i^0 = 0 \quad \forall i = 1, \dots, n,$$

$$y_i = 0 \iff y_i^0 = 0 \quad \forall i = 1, \dots, n.$$

Consider now the following index sets

$$A := \{i | x_i \neq 0, y_i \neq 0\}, \quad B := \{i | x_i = 0, y_i \neq 0\}, \quad C := \{i | x_i \neq 0, y_i = 0\},$$

and then the function $\tilde{F} : (0, 1)^{s+t} \rightarrow \mathbb{R}$ given by

$$\tilde{F}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) := \sum_{i \in A} \left(x_i \ln \frac{x_i}{x_i + y_i} + y_i \ln \frac{y_i}{x_i + y_i} \right) - \sum_{i \in A \cup B} x_i \ln p_i - \sum_{i \in A \cup B} y_i \ln(1 - p_i),$$

where $\tilde{\mathbf{x}} = (x_{i_1}, \dots, x_{i_s})$ is the vector of positive components of \mathbf{x} and $\tilde{\mathbf{y}} = (y_{i_1}, \dots, y_{i_t})$ is the vector of positive components of \mathbf{y} .

Note that by the definitions (17) and (18) it follows that $\Delta(\mathbf{p}, \alpha) \equiv \tilde{F}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ and then $\tilde{F}(\tilde{\mathbf{x}}^0, \tilde{\mathbf{y}}^0) = \tilde{F}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$. Moreover $\Delta(\mathbf{p}, \cdot)$ is constant in the segment $[\alpha, \alpha^0]$ (the set of minimal point is convex) and then \tilde{F} is constant in the direction $(\tilde{\mathbf{x}} - \tilde{\mathbf{x}}^0, \tilde{\mathbf{y}} - \tilde{\mathbf{y}}^0)$. So we have

$$\nabla \tilde{F}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \cdot (\tilde{\mathbf{x}} - \tilde{\mathbf{x}}^0, \tilde{\mathbf{y}} - \tilde{\mathbf{y}}^0) + \tilde{F}(\tilde{\mathbf{x}}^0, \tilde{\mathbf{y}}^0) = \tilde{F}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}), \quad (19)$$

as $\nabla \tilde{F}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \cdot (\tilde{\mathbf{x}} - \tilde{\mathbf{x}}^0, \tilde{\mathbf{y}} - \tilde{\mathbf{y}}^0)$ is the directional derivative of \tilde{F} in the direction $(\tilde{\mathbf{x}} - \tilde{\mathbf{x}}^0, \tilde{\mathbf{y}} - \tilde{\mathbf{y}}^0)$.

Since

$$\nabla \tilde{F}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = \left(\sum_{i \in A} \ln \frac{x_i}{x_i + y_i} - \sum_{i \in A \cup B} \ln p_i, \sum_{i \in A} \ln \frac{y_i}{x_i + y_i} - \sum_{i \in A \cup B} \ln(1 - p_i) \right),$$

the condition (19) becomes

$$\sum_{i \in A} x_i \ln \frac{x_i^0(x_i + y_i)}{x_i(x_i^0 + y_i^0)} + y_i \ln \frac{y_i^0(x_i + y_i)}{y_i(x_i^0 + y_i^0)} = 0 \iff \sum_{i \in A} x_i \ln \frac{x_i^0(x_i + y_i)}{x_i(x_i^0 + y_i^0)} + \frac{x_i y_i^0 - x_i^0 y_i}{x_i^0 + y_i^0} + y_i \ln \frac{y_i^0(x_i + y_i)}{y_i(x_i^0 + y_i^0)} + \frac{y_i x_i^0 - y_i^0 x_i}{x_i^0 + y_i^0} = 0.$$

Every term in the previous sum can be reduced in two terms of the form

$$\ln t - t + 1, \quad t \in (0, +\infty),$$

and by (12) we know that each one of such terms is nonpositive and it vanishes if and only if $t = 1$. Therefore by Eq. (19) it follows

$$\frac{x_i^0(x_i + y_i)}{x_i(x_i^0 + y_i^0)} = 1 \quad \forall i \in A,$$

i.e.

$$\frac{x_i^0}{x_i^0 + y_i^0} = \frac{x_i}{x_i + y_i} \quad \forall i \in A,$$

that is

$$\frac{\alpha^0(E_i H_i)}{\alpha^0(E_i H_i) + \alpha^0(E_i^c H_i)} = \frac{\alpha(E_i H_i)}{\alpha(E_i H_i) + \alpha(E_i^c H_i)} \quad \forall i \in A.$$

It is easy to see that for the index $i \notin A$ we have $(\mathbf{q}_\alpha)_i = (\mathbf{q}_{\alpha^0})_i = 0$ or $(\mathbf{q}_\alpha)_i = (\mathbf{q}_{\alpha^0})_i = 1$. So we have proved that $\mathbf{q}_\alpha \equiv \mathbf{q}_{\alpha^0}$.

Finally note that we can suppose that (15) and (16) hold without loss of generality. In fact, every distribution β in the segment (α, α^0) minimizes $\Delta(\mathbf{p}, \cdot)$ and gives

$$(\mathbf{q}_\beta)_i = 0 \iff (\mathbf{q}_\alpha)_i = (\mathbf{q}_{\alpha^0})_i = 0 \quad \forall i = 1, \dots, n,$$

$$(\mathbf{q}_\beta)_i = 1 \iff (\mathbf{q}_\alpha)_i = (\mathbf{q}_{\alpha^0})_i = 1 \quad \forall i = 1, \dots, n.$$

Hence all the distributions in (α, α^0) generate assessments with 0 or 1 in the same components and then give the same correction.

Since \mathbf{q} is continuous on \mathcal{A}_1 we have $\mathbf{q}_\alpha \equiv \mathbf{q}_{\alpha^0}$ and then

$$(\mathbf{q}_\alpha)_i = 0 \iff (\mathbf{q}_{\alpha^0})_i = 0 \quad \forall i = 1, \dots, n,$$

$$(\mathbf{q}_\alpha)_i = 1 \iff (\mathbf{q}_{\alpha^0})_i = 1 \quad \forall i = 1, \dots, n. \quad \square$$

Separate consideration is needed when the minimum of $\Delta(\mathbf{p}, \cdot)$ is attained in $\mathcal{A}_0 \setminus \mathcal{A}_1$. In fact in such a case the distribution that minimizes induces some $\alpha(H_i) = 0$.

Theorem 4. If $\Delta(\mathbf{p}, \cdot)$ attains its minimum value on $\mathcal{A}_0 \setminus \mathcal{A}_1$ and $\alpha, \alpha^0 \in \mathcal{A}_0 \setminus \mathcal{A}_1$ are distributions that minimize $\Delta(\mathbf{p}, \cdot)$ with the same null conditioning events, then $(\mathbf{q}_\alpha)_i = (\mathbf{q}_{\alpha^0})_i$ for all i such that $\alpha(H_i) > 0$.

Proof. Without loss of generality we assume that $\alpha(H_i) = \alpha^0(H_i) = 0$, $i = 1, \dots, t$, while $\alpha(H_l) > 0$ and $\alpha^0(H_l) > 0$, $l = t + 1, \dots, n$.

Let $\bar{\mathcal{E}} = \{E_{t+1}|H_{t+1}, \dots, E_n|H_n\}$, $\bar{\mathbf{p}} = (p_{t+1}, \dots, p_n)$ and $\bar{\Omega}$ the sample space spanned by $\bar{\mathcal{E}}$. Notice that the elements of $\bar{\Omega}$ are pairwise disjoint and are unions of elements of Ω .

Consider the distributions β and β^0 over $\bar{\Omega}$ defined by

$$\beta_j = \sum_{\omega_i \subset \bar{\omega}_j} \alpha_i \quad \forall \bar{\omega}_j \in \bar{\Omega}; \quad (20)$$

$$\beta_j^0 = \sum_{\omega_i \subset \bar{\omega}_j} \alpha_i^0 \quad \forall \bar{\omega}_j \in \bar{\Omega}. \quad (21)$$

It is easy to see that

$$\beta(H_l) = \alpha(H_l) > 0, \quad (\mathbf{q}_\alpha)_l = (\mathbf{q}_\beta)_l = q_{\beta_l}, \quad l = t + 1, \dots, n, \quad (22)$$

$$\beta^0(H_l) = \alpha^0(H_l) > 0, \quad (\mathbf{q}_{\alpha^0})_l = (\mathbf{q}_{\beta^0})_l = q_{\beta_l^0}, \quad l = t + 1, \dots, n, \quad (23)$$

so that both β and β^0 result in $\bar{\mathcal{A}}_1$ (the analogous of \mathcal{A}_1 associated to $\bar{\mathcal{E}}$).

Moreover

$$\Delta(\bar{\mathbf{p}}, \beta) = \sum_{l=t+1}^n \beta(H_l) \left[q_{\beta_l} \ln \left(\frac{q_{\beta_l}}{p_l} \right) + (1 - q_{\beta_l}) \ln \left(\frac{1 - q_{\beta_l}}{1 - p_l} \right) \right] = \Delta(\mathbf{p}, \alpha),$$

and analogously we have

$$\Delta(\bar{\mathbf{p}}, \beta^0) = \Delta(\mathbf{p}, \alpha^0).$$

Since by hypothesis $\Delta(\mathbf{p}, \alpha) = \Delta(\mathbf{p}, \alpha^0)$, we have $\Delta(\bar{\mathbf{p}}, \beta) = \Delta(\bar{\mathbf{p}}, \beta^0)$ and applying [Theorem 3](#) on $\bar{\Omega}$ it follows $\mathbf{q}_\beta = \mathbf{q}_{\beta^0}$; hence $(\mathbf{q}_\alpha)_l = (\mathbf{q}_{\alpha^0})_l$, $l = t + 1, \dots, n$. \square

Theorem 5. *If $\alpha, \alpha^0 \in \mathcal{A}_0$ are distributions that minimize $\Delta(\mathbf{p}, \cdot)$, then for all $i \in \{1, \dots, n\}$ with $\alpha(H_i) > 0$ and $\alpha^0(H_i) > 0$ we have $(\mathbf{q}_\alpha)_i = (\mathbf{q}_{\alpha^0})_i$.*

Proof. Let I and I^0 subsets of $\{1, \dots, n\}$ defined by

$$I = \{i | \alpha(H_i) > 0\}, \quad I^0 = \{i | \alpha^0(H_i) > 0\}.$$

It is easy to see that all the probability distributions

$$\delta_r = r\alpha + (1-r)\alpha^0, \quad r \in (0, 1),$$

minimize $\Delta(\mathbf{p}, \cdot)$ and $\delta_r(H_i) > 0$ for $i \in I \cap I^0$.

Therefore, by [Theorem 4](#) it follows that for $i \in I \cap I^0$

$$(\mathbf{q}_{\delta_r})_i = (\mathbf{q}_{\delta_s})_i \quad \forall r, s \in (0, 1),$$

and then, by continuity

$$(\mathbf{q}_\alpha)_i = (\mathbf{q}_{\alpha^0})_i \quad \forall i \in I \cap I^0. \quad \square$$

Hence, if $\alpha \in \mathcal{A}_0$ minimizes $\Delta(\mathbf{p}, \cdot)$ then for all $i \in \{1, \dots, n\}$ such that $\alpha(H_i) > 0$ we can assert that \mathbf{q}_i is properly determined.

Observe that [Theorem 5](#) is a generalization of [Theorems 3 and 4](#). In fact, if $I = I^0 = \{1, \dots, n\}$ the hypotheses of [Theorem 3](#) are satisfied, while if $I = I^0 \neq \{1, \dots, n\}$ [Theorem 4](#) applies.

Theorem 6. *There exists $\tilde{\alpha}$ in \mathcal{A}_0 such that $\Delta(\mathbf{p}, \cdot)$ attains its minimum on $\tilde{\alpha}$ and such that the number of H_i with $\tilde{\alpha}(H_i) > 0$ is maximum. Moreover the associated $\tilde{\mathbf{q}}_i$ are uniquely determined.*

Proof. Let

$$A := \{\alpha | \Delta(\mathbf{p}, \alpha) \text{ is minimum}\},$$

and

$$J := \{i : \exists \alpha \in A \text{ with } \alpha(H_i) > 0\}.$$

For any $j \in J$ choose an $\alpha^j \in A$ such that $\alpha^j(H_j) > 0$. Every $\tilde{\alpha}$ (proper) convex combination of $\{\alpha^j, j \in J\}$ verifies the statement. \square

As a straightforward consequence, by definition of \mathcal{A}_1 , we have the following:

Corollary 7. *If there exists $\alpha \in \mathcal{A}_1$ such that $\Delta(\mathbf{p}, \cdot)$ attains its minimum on α , then all the q_i are uniquely determined as $(\mathbf{q}_\alpha)_i$, $i = 1, \dots, n$.*

It is quite common to find a minimizer α in \mathcal{A}_1 . For example, whenever \mathbf{p} is unconditional then, obviously, all the distributions which minimize Δ are in \mathcal{A}_1 . With respect to a general conditional framework, we found e.g. that all the examples given in [\[2\]](#) express a minimum in \mathcal{A}_1 . Anyhow it is hard to characterize, from a theoretical point of view, the initial assessments $(\mathcal{E}, \mathbf{p})$ which produce a minimum in \mathcal{A}_1 . We can guess that this mainly depends on the logical structure of \mathcal{E} similarly to what happens for the so called *locally strong coherence* (see [\[4\]](#)). On the other hand, from a practical point of view, the procedure suggested in [Section 4](#) suddenly reveals if there is a minimum in \mathcal{A}_1 in its second step.

Note that [Theorems 3 and 4](#), apart from the existence of the optimal solution they guarantee, are quite technical. Nevertheless [Theorem 6](#) and its subsequent [Corollary 7](#) enlighten their significance from a practical point of view and they suggest us how to proceed to find the solution, as it will be described in the next section.

4. Our correction procedure

The formal properties of the discrepancy $\Delta(\mathbf{p}, \alpha)$ proved in the previous section permit us to propose a procedure to adjust an initial incoherent assessment with a coherent one.

The initial problem is to find an α in \mathcal{A}_0 with minimum $\Delta(\mathbf{p}, \alpha)$. This procedure starts by looking for a solution α having maximum \mathcal{E}_α (whose existence is guaranteed by [Theorem 6](#)). In fact step (i2) looks for a solution α which maximizes the number of H_i such that $\alpha(H_i) > 0$. Moreover, theorems in previous section ensure the existence of a unique coherent conditional assessment \mathbf{q}_α “close” as much as possible to \mathbf{p} . Unluckily such assessment \mathbf{q}_α could not be defined overall \mathcal{E} but only

on \mathcal{E}_x . This is because there could be some conditional events $E_i|H_i$ -s such that the H_i -s have null probability for all the distributions that induce \mathbf{q}_x . The further steps of the procedure extend \mathbf{q}_x to the rest of \mathcal{E} so that $\Delta(\mathbf{p}, \alpha)$ remains minimum.

Coherence for the correction \mathbf{q}_x requires that

$$\alpha(H_i) = (\mathbf{q}_x)_i \alpha(E_i|H_i), \quad i = 1, \dots, n, \quad (24)$$

and suppose the first t constraints are verified by any value of $(\mathbf{q}_x)_i$ because $\alpha(H_i) = 0$, $i = 1, \dots, t$. For the overall correction we could take as good only the components $(\mathbf{q}_x)_j$, $j = t + 1, \dots, n$. For the first t components we could focus our attention on the restriction of the assessment $\mathbf{p}^2 = (p_1, \dots, p_t)$ that could be seen as an initial assessment on the sub-domain $\mathcal{E}^2 = \{E_1|H_1, \dots, E_t|H_t\}$. If such assessment \mathbf{p}^2 results by itself incoherent, we could correct it with the same methodology illustrated before, obtaining a $\mathbf{q}_{x^2} = ((\mathbf{q}_{x^2})_1, \dots, (\mathbf{q}_{x^2})_t)$ from which to select only the significant components, i.e. those $(\mathbf{q}_{x^2})_j$ with associated conditioning event H_j with positive probability $\alpha^2(H_j)$. This procedure can be iterated, at worst n times, with uniqueness of the significant components at each step. Obviously, whenever the conditions of Corollary 7 are fulfilled, i.e. whenever there exists a minimizing distribution α such that $\alpha(H_i) > 0$ for all i , the procedure will produce the unique coherent correction \mathbf{q}_x through a single iteration. This can happen for example if we have an unconditional initial assessment \mathbf{p} .

Let us describe the full procedure in details.

Let \mathcal{E} , \mathbf{p} , Ω , \mathcal{A}_0 be defined as in Section 2, then:

(i0) Set

$$l := 1,$$

$$\mathcal{E}^l := \mathcal{E},$$

$$\mathbf{p}^l := \mathbf{p},$$

$$\Omega^l := \Omega,$$

$$\mathcal{A}_0^l := \mathcal{A}_0;$$

(i1) Let α^l be a solution of the nonlinear optimization program

$$\min \quad \Delta(\mathbf{p}^l, \alpha^l) \quad (25)$$

$$\text{s.t.} \quad \alpha^l \in \mathcal{A}_0^l; \quad (26)$$

(i2) Let α^l be a solution of the mixed integer program

$$\max \quad \sum_{E_j|H_j \in \mathcal{E}^l} I(\alpha^l(H_j)) \quad (27)$$

$$\text{s.t.} \quad \alpha^l \in \mathcal{A}_0^l, \quad (28)$$

$$\Delta(\mathbf{p}^l, \alpha^l) = \Delta(\mathbf{p}^l, \tilde{\alpha}^l), \quad (29)$$

$$I(\alpha^l(H_i)) = \begin{cases} 1 & \text{if } \alpha^l(H_i) > 0, \\ 0 & \text{if } \alpha^l(H_i) = 0. \end{cases} \quad (30)$$

(i3) For any $E_j|H_j \in \mathcal{E}^l$, if $\tilde{\alpha}^l(H_j) > 0$ then set

$$(\mathbf{q}_{\alpha^l})_j := \frac{\tilde{\alpha}^l(E_j|H_j)}{\tilde{\alpha}^l(H_j)}; \quad (31)$$

(i4) Set

$$\mathcal{E}^{l+1} := \{E_j|H_j \in \mathcal{E}^l : \tilde{\alpha}^l(H_j) = 0\},$$

$$\mathbf{p}^{l+1} := (p_j \in \mathbf{p}^l : \tilde{\alpha}^l(H_j) = 0),$$

$$\Omega^{l+1} := \{\omega_w \in \Omega^l : \tilde{\alpha}^l_w = 0\},$$

$$\mathcal{A}_0^{l+1} := \left\{ \alpha' \in \mathcal{A}_0^l : \alpha' \left(\bigvee_{E_j|H_j \in \mathcal{E}^l} H_j \right) = 1 \right\},$$

$$l := l + 1;$$

(i5) If $\mathcal{E}^l \neq \emptyset$ then return to (i1) else exit.

As coherent correction of \mathbf{p} we get the assessment \mathbf{q}_x generated by the distributions $\tilde{\alpha}_1, \dots, \tilde{\alpha}_m$ through (31).

Since the various $\tilde{\alpha}_i$'s are probability distributions over the respective \mathcal{A}_0^l and the $(\mathbf{q}_x)_j$, $j = 1, \dots, n$, are obtained through (31), coherence of the overall \mathbf{q}_x derives from theorem of Coletti in [5] that we report here for completeness:

Theorem 8. Let $\mathcal{E} = [E_1|H_1, \dots, E_n|H_n]$ be an arbitrary finite family of conditional events and let Ω^0 denote the set of atoms ω_j generated by the events $E_1, H_1, \dots, E_n, H_n$. A real function \mathbf{q} is a coherent conditional probability assessment on \mathcal{E} if and only if there exists (at least) a class of probabilities $\{\alpha^0, \dots, \alpha^t\}$ such that each α^l is defined on a subset $\Omega^l \subseteq \Omega^0$ and for any $E_i|H_i \in \mathcal{E}$ there is a unique α^l with

$$\sum_{\omega_j \subseteq H_i} \alpha_j^l > 0, \quad q(E_i|H_i) = \frac{\sum_{\omega_j \subseteq E_i H_i} \alpha_j^l}{\sum_{\omega_j \subseteq H_i} \alpha_j^l}, \quad (32)$$

and $\Omega^{l'} \subset \Omega^{l''}$ for $l' > l''$, $\alpha_j^{l''} = 0$ if $\omega_j \in \Omega^{l'}$.

Note that, by Theorem 6, the objective function (27) of the mixed integer programs in (i2) guarantees the identification of a maximum number of significant components for \mathbf{q}_x . The constraint (29) restricts the search of the optimal solution amongst those that minimize the discrepancy $\Delta(\mathbf{p}, \cdot)$, as determined in (i1). Operationally, such mixed integer programs (27)–(30) can be solved through equivalent ordinary nonlinear programs.

Note moreover that the optimal solutions $\hat{\alpha}^l$ and $\hat{\alpha}^l$ of the optimization problems in (i1) and (i2) are auxiliary components. Actually, our attention is focused on \mathbf{q}_x as “best” approximation of \mathbf{p} , that implicitly admits as reasonable models not only the $\hat{\alpha}^l$ ’s that had generated it, but also all other agreeing distributions.

Obviously, if the initial assessment \mathbf{p} is coherent “per se” overall \mathcal{E} but involving events H_i with compulsory null probability, the procedure generates a coincident assessment $\mathbf{q}_x \equiv \mathbf{p}$. In fact any sub-assessment \mathbf{p}^l results a coherent assessment over Ω^l , hence the nonlinear problems in steps (i1) return optimal values $\hat{\alpha}^l$ such that $\Delta(\mathbf{p}^l, \hat{\alpha}^l) = 0$. Consequently constraints (29) guarantee that also the optimal solutions $\hat{\alpha}^l$ of steps (i2) have $\Delta(\mathbf{p}^l, \hat{\alpha}^l) = 0$ and hence, by Theorem 1, condition (ii), we get $\mathbf{q}_x \equiv \mathbf{p}^l$. The fact that $\hat{\alpha}^l$ ’s are optimal solutions of the mixed integer programs (27)–(30) implies the use of the least number of zero-layers in the reconstruction of \mathbf{p} .

A direct implementation of our procedure could incur in complexity troubles. The verification of coherence is already a NP-complete problem “per se”. As a consequence, our nonlinear optimization problems (i1) and (i2) are even harder. Modern optimization tools like GAMS [1] make medium-size problems treatable with some tens of events. The main dimensional problem is that an explicit expression of the unknowns α^l require an $O(3^n)$ number of components. In fact, in the worst case of absence of logical constraints among the E_i ’s and the H_i ’s, the number of unknowns needed to perform the optimization problems is $3^n - 1$. Number that reduces to 2^n in the special case of an unconditional assessment. But the worst case study is meaningless for our purpose because any assessment results coherent “per se”. The actual number of atoms in Ω is usually smaller because the presence of logical constraints among the basic events E_i ’s and H_i ’s, and in particular the more constraints are present the less atoms we have. This of course does not solve the dimensional explosion in large-size problems, where implicit expression of the solution and heuristics are needed.

Note finally that if we deal with an unconditional assessment \mathbf{p} , apart from the relative small benefit in the dimensional complexity outlined before, we are only assured that all the distributions α that minimize $\Delta(\mathbf{p}, \cdot)$ lie in \mathcal{A}_1 so that it will be needed only one optimization step (i1) to get a distribution $\hat{\alpha}$ and consequently the full correction \mathbf{q}_x . The nonlinear nature of the optimization problem (i1) anyway does not change since the presence of the log’s in the expression of $\Delta(\mathbf{p}, \alpha)$.

5. Conclusions

With this contribution we intended to give a detailed exposition of the formal properties of the discrepancy $\Delta(\mathbf{p}, \alpha)$. Thanks in fact to its peculiarities, summarized by the results of the theorems in Section 3, the correction procedure proposed in Section 4 acquires significance and operational soundness.

The discrepancy measure $\Delta(\mathbf{p}, \alpha)$ revealed a nice tool not only for the correction procedure, as shown by prototypical examples in [2], but also to merge different expert opinions in both precise and imprecise probability contexts, as depicted in [3]. In the latter contribution our approach has been generalized to the imprecise framework by considering several sets of assessments to approximate by single coherent assessments. In fact we consider the set of assessments $\mathcal{V} = \{\mathbf{v}\}$ obtained by fixing one range bound and letting all other components vary inside the remaining ranges. We can find a coherent approximation \mathbf{p}_x of $\{\mathbf{v}\}$ by minimizing $\Delta(\mathbf{v}, \alpha)$, with $\mathbf{v} \in \mathcal{V}$. This process is iterated by varying the fixed bound, obtaining at the end a set of coherent assessments whose lower/upper envelope can be adopted as correction of the initial (incoherent) ranges. At the moment there is an open problem about the uniqueness of each single approximation \mathbf{p}_x . Empirical studies have corroborated the conjecture of uniqueness, but a formal proof is still missing.

By referring to imprecise probabilities a comparison with Walley’s theory [16] is compulsory. It is known that in Walley’s behavioral theory, different notions of conditional coherence are possible. As notably remarked in [13], the different notions can be mainly interpreted in a Bayesian sensitivity style through classes of agreeing (i.e. with values inside the ranges) precise conditional assessments, so that we can use this formulation to compare his approach with our. A detailed comparison is outside the scope of the present contribution. Anyhow we can assert that: if the initial assessment avoids partial loss, i.e. the class of agreeing precise coherent conditional assessments is not empty, our procedure should produce a correction that coincides with Walley’s notion of natural extension. In particular, if the initial assessment is coherent in Walley’s sense, i.e. it coincides with the envelope of the agreeing precise coherent conditional assessments, the correction procedure do not modify it. About Walley’s weak coherence, as proved in [13], its distinction with coherence appears whenever some event is conditioned to zero probabilities. This means that if we start with an initial assessment that is weakly coherent but not

coherent, our correction procedure will not modify the elements whose conditioning event can have positive probability, while it will change only those that are obliged to have zero-probability. What can be remarked here is that our procedure acquires relevance in presence of an initial assessment that *incurs in a partial loss*, i.e. whenever the set of agreeing precise coherent assessments is empty. In such situation the natural extension procedure cannot be applied, while ours can proceed undisturbed.

Of course, our proposal is not a *panacea* to solve incoherence for conditional probability assessments. But the behavioral origin of the discrepancy through the scoring rule (see Eq. (6)) and its useful decomposition on the separate scenarios (see Eq. (9)) should be enough to justify its adoption. In fact in [2] we have also made some numerical comparisons among corrections obtained through other divergencies. Those that performed similarly to our discrepancy $\Delta(\mathbf{p}, \alpha)$ were those with only geometrical motivations, like usual $L1$ and $L2$ metric distances suggested in [15], but without either an intuitive or a probabilistic interpretation to be used as distances between conditional assessments. A straightforward adaptation of the usual logarithmic Bregman divergence to conditional probabilities produced less meaningful results, especially in presence of “heavy” incoherences in the initial assessment. This is mainly because in its generating logarithmic scoring rule only the events that occur are taken, without considering those that turn out to be false. On the other side, others more intuitive divergences, that could maintain the relative proportions among the components p_i ’s, or among components p_i ’s and their complements $1 - p_i$, expressed computational drawbacks due to the presence of local minima.

The main development of our work will be to find practical applications where to test the effective goodness of the results and the effective solvability of the optimization steps (i1) and (i2) with large scale domains.

References

- [1] A. Brooke, D. Kendrick, A. Meeraus, R. Raman, GAMS: A Users Guide, GAMS Development Corp., Washington, DC, 2003.
- [2] A. Capotorti, G. Regoli, Coherent correction of inconsistent conditional probability assessments, in: Proceedings of IPMU’08 – Malaga (Es), 2008, pp. 891–898.
- [3] A. Capotorti, G. Regoli, F. Vattari, On the use of a new discrepancy measure to correct incoherent assessments and to aggregate conflicting opinions based on imprecise conditional probabilities, in: Proceedings of ISIPTA’09 International Conference, 2009, pp. 71–78.
- [4] A. Capotorti, B. Vantaggi, Locally strong coherence in an inference process, *Annals of Mathematics and Artificial Intelligence* 35 (1–4) (2002) 125–149.
- [5] G. Coletti, Coherent numerical and ordinal probabilistic assessments, *IEEE Transaction on Systems, Man, and Cybernetics* 24 (1994) 1747–1754.
- [6] G. Coletti, R. Scozzafava, Probabilistic Logic in a Coherent Setting, *Trends in Logic*, Kluwer, Dordrecht, 2002.
- [7] B. de Finetti, Sull’impostazione assiomatica del calcolo delle probabilità, *Annali Univ. Trieste* 19 (1949) 29–81. Engl. transl. of *Probability, Induction, Statistics* Ch.5. Wiley, London, 1972.
- [8] L.E. Dubins, Finitely additive conditional probabilities, conglomerability and disintegration, *The Annals of Probability* 3 (1975) 89–99.
- [9] S. Holzer, On coherence and conditional prevision, *Bull. Unione Matematica Italiana, Analisi funzionale e applicazioni* 6 (4) (1985) 441–460.
- [10] O. Kriz, Conditional problem for objective probability, *Kybernetika* 34 (1) (1998) 27–40.
- [11] P.H. Krauss, Representation of conditional probability measures on Boolean algebras, *Acta Mathematica Academiae Scientiarum Hungaricae* 19 (1968) 229–241.
- [12] F. Lad, *Operational Subjective Statistical Methods: A Mathematical, Philosophical, and Historical Introduction*, John Wiley, New York, 1996.
- [13] E. Miranda, Updating coherent previsions on finite spaces, *Fuzzy Sets and Systems* 160 (2009) 1286–1307.
- [14] A. Rényi, On a new axiomatic theory of probability, *Acta Mathematica Academiae Scientiarum Hungaricae* 6 (1955) 285–335.
- [15] B. Vantaggi, Statistical matching of multiple sources: a look through coherence, *International Journal of Approximate Reasoning* 49 (3) (2008) 701–711.
- [16] P. Walley, *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, London, 1991.